Multiple corresponding analysis with missing data

Peter G.M. van Der Heijden, Brigitte Escofier

INRIA

UNITÉ DE RECHERCHE
INRIA-RENNES

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
B.P 105
78153 Le Chesnay Cedex
France
Tél. (1) 39 63 55 11

Rapports de Recherche

N° 902

MULTIPLE CORRESPONDING ANALYSIS WITH MISSING DATA

Programme 5

Peter G. M. van der HEIJDEN
Brigitte ESCOFIER

Septembre 1988

RR-0902

MULTIPLE CORRESPONDING ANALYSIS

WITH MISSING DATA

Peter G.M. van der HEIJDEN
Brigitte ESCOFIER

Septembre 1988

Publication Interne n° 423

IRISA

INSTITUT DE RECHERCHE EN INFORMATIQUE
ET SYSTÈMES ALÉATOIRES

Campus Universitaire de Beaulieu
35042 - RENNES CÉDEX
FRANCE
Téléphone : 99 36 20 00
Télex : UNIRISA 950 473 F
Télécopie : 99 38 38 32

Multiple correspondence analysis with missing data

Peter G.M. van der Heijden* & Brigitte Escofier**

Publication Interne n° 423
Septembre 1988
30 Pages

Summary

For multiple correspondence analysis many procedures are known for handling missing data. In this paper some properties of the most important procedures are discussed. As a second step distinct types of missing data are distinguished. Apart from a typology presented by Little & Rubin (1987), we also give attention to the fact that in multiple correspondence analysis regularly missing data are created on purpose in order to eliminate the influence of specific categories. Given the properties of the different procedures for handling missing data, it is discussed which types of missing data can best be handled with which procedures for missing data.

This paper was written while the first author visited I.R.I.S.A. in Rennes. We are grateful to S. Nishisato and J.J.A. van Rijckevorsel for helpful comments on an earlier version of this paper. *Department of Psychometrics and Research Methods, University of Leiden, The Netherlands. **I.R.I.S.A., Rennes, France.

Traitement des valeurs manquantes en analyse des correspondances multiples

Peter G.M. van der Heijden* & Brigitte Escofier**

Résumé

Beaucoup de méthodes ont été proposées pour traiter les valeurs manquantes en analyse des correspondances multiples. Dans cet article, on étudie les propriété des méthodes les plus importantes. Nous distinguons différents types de valeurs manquantes en suivant des propostitions de Little & Rubin (1987), et noud étudions aussi le cas de valeurs manquantes crées pour éliminer l'influence de catégories particulières. Les propriétés des differentes méthodes sont utilisées pour étudier celles qui s'adaptent le mieux à chaque type de valeurs manquantes.

This paper was written while the first author visited I.R.I.S.A. in Rennes. We are grateful to S. Nishisato and J.J.A. van Rijckevorsel for helpful comments on an earlier version of this paper. *Department of Psychometrics and Research Methods, University of Leiden, The Netherlands. **I.R.I.S.A., Rennes, France.

1. Introduction

Multiple correspondence analysis (MCA) is a tool for data description that receives much attention in the last decade. Many reasons for this can be given, one of it being that it is a very flexible tool in the sense that it can be applied to many types of data (see for examples Nishisato, 1980; Gifi, 1981; Greenacre, 1984). In this paper we discuss MCA of categorical data that are coded into an object by variable matrix in some way or another. Such data are very common in the social sciences, for example as a result of using questionaires. We concentrate on the case that the object by variable matrix contains missing values.

For object by variable data many different missing data procedures exist in MCA. The major sources for these procedures are Hamrouni & Benzécri (1976), Nishisato (1980), Meulman (1982), Greenacre (1984) and Benali (1985). In this paper we give an overview of the most important missing data procedures. In our opinion this overview is much needed, since the above authors often do not seem to be aware of each others work. We will compare these procedures systematically by studying their properties.

Using a typology of missing data provided by Little & Rubin (1987). They distinguish in the first place data that are really missing from data that are not really missing. An example of the latter class is that objects cannot be classified in the original categories, for example, when a question is irrelevant for some persons. As an extra type we define missing data that are created on purpose, for example, in order to eliminate the influence of diminating categories. We will try to come to conclusions that in some specific missing data situation some specific procedure seems better suited than another. We start with a description of MCA for non-missing data.

2. Multiple correspondence analysis

There are many ways to introduce MCA. We will introduce MCA here shortly as simple correspondence analysis (CA) of an indicator matrix (see, for example, Greenacre, 1984, ch. 5). Many other approaches to MCA are possible, see, for example, Nishisato (1980), Gifi (1981), Lebart, Morineau & Warwick (1984). For a summary we refer to Tenenhaus & Young (1985). Introductions to simple CA are also provided by the above authors. We will emphasize CA properties that result from the fact that an indicator matrix is analyzed. We only discuss properties needed for our discussion of the various approaches to missing values in MCA: this paper is not meant as an introduction to MCA. For more details we refer to the authors mentioned above.

An indicator matrix is defined in the following way. Consider a data matrix P with qualitative measures of n objects indexed by i, on m variables indexed by j. See table 1a. Each variable j has k_j categories, indexed by l, and the total number of categories is $k=\Sigma_j k_j$. We transform the matrix P into a binary *indicator matrix* X of order n x k, where each category l of variable j has its own column. In this matrix $x_{ijl} = 1$ if object i falls into category l of variable j, and $x_{ijl} = 0$ else. See table 1b. Notice that $x_{ij+}=1$, hence $x_{i++}=m$ and $x_{+++}=nm$, and x_{+jl} is the marginal frequency for category l of variable j. We will now discuss CA of this matrix, emphasizing geometrical properties (see also Meulman, 1982, 1986; Greenacre 1984; Carroll, Green & Schaffer, 1986).

CA gives two geometrical representations of a matrix, one for the rows and one for the columns. First consider the representation for the rows. Each row is represented by a point in k-dimensional *euclidean space*, and has coordinates equal to x_{ijl}/x_{i++}. The metric in this space is defined by D_c^{-1}, where D_c is diagonal with elements the column margins $x_{+jl}/x_{+++} = x_{+jl}/nm$. The vector of coordinates x_{ijl}/x_{i++} is defined as the *profile* for row i. With each point i a mass $x_{i++}/x_{+++} = 1/n$ is associated. The distance between the profiles for row i and i' is called the *chi-squared distance*, and is defined as

$$\delta^2(i,i') = \Sigma_{jl} (x_{+++}/x_{+jl}) (x_{ijl}/x_{i++} - x_{i'jl}/x_{i'++})^2 \qquad (1a)$$

Since $x_{i++}=x_{i'++}=m$ and $x_{+++}=nm$, (1a) can be simplified to

$$\delta^2(i,i') = (n/m)\ \Sigma_{jl}\ (1/x_{+jl})\ (x_{ijl} - x_{i'jl})^2 \qquad (1b)$$

This shows that combination jl does not increase $\delta^2(i,i')$ when objects i and i' either both fall or both do not fall into the same category l of variable j: in both cases $x_{ijl}-x_{i'jl}=0$. When only one of the two objects falls into jl, the increase of $\delta^2(i,i')$ for jl is proportional to $(1/x_{+jl})$. Therefore objects i and i' will be close together when they have many categories in common; this also shows that $\delta^2(i,i')$ will become larger when i and i' differ by falling into distinct categories with lower marginal frequencies.

We will now study the distance of object i to the average row profile O, which is the profile of the column margins. This average row profile is in the weighted average of all row profiles, using masses x_{i++}/x_{+++} as weights, and has values x_{+jl}/x_{+++}. So using (1a) $\delta^2(i,O)$ can be found to be

$$\delta^2(i,O) = (n/m)\ \Sigma_{jl}\ (1/x_{+jl})\ (x_{ijl} - (x_{+jl}/n))^2 \qquad (1c)$$

This shows that when object i does not fall into jl, then $x_{ijl}=0$, and $\delta^2(i,O)$ increases with x_{+jl}/nm for column jl; when object i falls into jl, then $x_{ijl}=1$, and $\delta^2(i,O)$ increases with $(n/mx_{+jl}) - (2/m) + (x_{+jl}/nm)$ for column jl. We conclude that when object i falls into categories with lower marginal frequencies x_{+jl}, $\delta^2(i,O)$ increases more.

We can also derive the weighted sum of squared distances of all objects to the origin. Since there are x_{+jl} objects falling into jl, and $(n-x_{+jl})$ objects not falling into jl, it follows that for jl the sum of all squared distances is $(n/m) - x_{+jl}/m$. Weighting this with the masses of the objects, and summing over all possible jl (which is allowed due to Pythagoras theorem), we find the so-called *total inertia*

$$\text{total inertia} = (k/m) - 1 \qquad (1d)$$

Now we discuss the chi-squared distances between the columns of X. Columns are represented as points in n-dimensional euclidean space, with coordinates for column jl x_{ijl}/x_{+jl}. The metric of the space is defined by the D_r^{-1}, with D_r diagonal having elements $x_{i++}/x_{+++}=1/n$, so we deal with a metric that is proportional to the identity metric. Each

column jl has a mass $x_{+jl}/x_{+++} = x_{+jl}/nm$ associated with it. The profile for category l of variable j has values x_{ijl}/x_{+jl}, and we find as the chi-squared distance between jl and j'l'

$$\delta^2(jl,j'l') = \Sigma_i \, (x_{+++}/x_{i++}) \, (x_{ijl}/x_{+jl} - x_{ij'l'}/x_{+j'l'})^2 \qquad (2a)$$

Since $x_{+++}=nm$ and $x_{i++}=m$, this simplifies to

$$\delta^2(jl,j'l') = n \, \Sigma_i \, (x_{ijl}/x_{+jl} - x_{ij'l'}/x_{+j'l'})^2 \qquad (2b)$$

If object i does not fall in either of x_{ijl} or $x_{ij'l'}$, then the increase of $\delta^2(jl,j'l')$ for object i is 0. If object i falls in only one of both categories, say in jl, then the increase in $\delta^2(jl,j'l')$ is $n(1/x_{+jl})^2$, which will be larger when jl has a lower marginal frequency. If object i falls in both jl and j'l', then the increase in $\delta^2(jl,j'l')$ is $n(1/x_{+jl} - 1/x_{+j'l'})^2$ which will be larger the more the marginal frequencies of jl and j'l' differ. Another way to show the influence of marginal frequencies on the distance between two columns is the following: if we work out (2b), using the fact that $x_{ijl}^2 = x_{ijl}$ and $x_{ij'l'}^2 = x_{ij'l'}$, we find

$$\delta^2(jl,j'l') = n \, (1 - \Sigma_i x_{ijl}x_{ij'l'})/x_{+jl} + n(1 - \Sigma_i x_{ijl}x_{ij'l'})/x_{+j'l'} \qquad (2c)$$

which shows that the distance between jl and j'l' is proportional to the sum of the proportion of objects having jl but not j'l' and the proportion of objects having j'l' but not jl. It also shows that the distance between two categories l and l' of the same variable j is proportional to $1/x_{+jl} + 1/x_{+jl'}$.

When we use (2a) to study the distance to the average column profile O, we find

$$\delta^2(jl,O) = n \, \Sigma_i \, (x_{ijl}/x_{+jl} - 1/n)^2 \qquad (2d)$$

This shows that, when object i does not fall into jl, then $x_{ijl}=0$, and $\delta^2(jl,O)$ increases with 1/n. When object i falls into jl, $x_{ijl}=1$, then $\delta^2(jl,O)$ increases with $(n/x_{+jl}^2) - (2/x_{+jl}) + (1/n)$. The total distance $\delta^2(jl,O)$ can easily be calculated as the sum of $(n-x_{+jk})/n$ (i.e. $n-x_{+jk}$ objects do not fall into jl) and x_{+jl} $((n/x_{+jl}^2) - (2/x_{+jl}) + (1/n))$ (since there are x_{+jl} objects falling into jl), and hence we find $\delta^2(jl,O) = (n/x_{+jl}) - 1$. This shows that the distance of jl to the average column profile is larger for a category with a smaller marginal

frequency. When we take the sum over all jl, weighted with the masses x_{+jl}/x_{++}, we find as total inertia (1d), showing that the total inertia of the configuration of row points is identical to the total inertia of the configuration of the column points.

We now give another way to derive the total inertia, that will be useful when we discuss the total inertia for different missing value approaches. This way is

$$\text{Total inertia} = \Sigma_i\Sigma_{jl} ((x_{ijl}/x_{+++}) - (x_{i++}x_{+jl}/x^2_{+++}))^2 / (x_{i++}/x_{+++})(x_{+jl}/x_{+++}) \qquad (3a)$$

and by working out the term in the numerator, using the fact that $x^2_{ijl} = x_{ijl}$, we find

$$\text{Total inertia} = (\Sigma_i\Sigma_{jl}\, x_{ijl}/(x_{i++}x_{+jl})) - 1 \qquad (3b)$$

and since $x_{i++} = m$, we can reduce (3b) to further to (1d).

Knowing the total inertia of the full-dimensional space, one is usually interested in the study of a low dimensional projection of this space that displays as much inertia of the full dimensional space as possible. One generalized singular value decomposition (SVD) can be used to project the full dimensional space for the rows and the full dimensional space for the columns onto lower dimensional subspaces (see, for example, Greenacre, 1984). This corresponds to the following decomposition of X:

$$X/x_{+++} = D_r(1 + R\Lambda C')D_c \qquad (4)$$

where D_r is diagonal with margins x_{i++}/x_{+++}, D_c is diagonal with margins x_{+jl}/x_{+++}, R is a matrix with object scores $r_{i\alpha}$ for row i on dimension α and C a matrix with category scores $c_{jl\alpha}$ for category l of variable j on dimension α, and Λ is a diagonal matrix with singular values λ_α. R is normalized so that $R'D_rR = I$ and $t'D_rR = 0$, and C is normalized similarly as $C'D_cC = I$ and $t'D_cC = 0$. Since $D_r = I/n$, and $x_{+++} = nm$, (4) can be simplified, and the normalization of R is $R'R = nI$ and $t'R = 0$.

The row scores R and column scores C are related by the transition formulas

$$R^* = R\Lambda = D_r^{-1}XC \quad (5a)$$

$$C^* = C\Lambda = D_c^{-1}X'R \quad (5b)$$

These properties are the key concepts in 'reciprocal averaging', one of the approaches to MCA. Since $D_r = I/n$, (5a) can be simplified. Equation (5a) shows that object scores can be derived as averages of the scores for the categories that they used, and (5b) shows that category scores can be derived as averages of the scores for the objects that used these categories. Distances between the rows are equal to chi-squared distances when R* is used as coordinates for the rows, and distances between the columns are equal to chi-squared distances when C* are used as coordinates for the columns. Often the pair (R,C*) is used for a joint representation in which the category points are in the weighted average of the objects falling into them. Notice that $R^*'D_rR^* = \Lambda^2 = C^*'D_cC^*$, where trace Λ^2 equals the total inertia. This shows that the total inertia is spread out over the row points and dimensions, and the column points and dimensions.

MCA is related to many techniques for the analysis of quantitative data, such as analysis of variance, principal components analysis and generalized canonical analysis. We will not discuss this in detail here, but refer instead to Tenenhaus & Young (1985). Here we will only discuss this relation verbally. Central in the relation between MCA and these techniques is the notion of the *quantified data matrix* that we will denote as Q. Starting from a non-quantified data matrix P (see table 1a), we can use the scores found for the categories in MCA to construct a quantified data matrix Q of the same order as P. In principle such a quantified data matrix can be constructed for each MCA dimension, but here we only describe this for the first dimension. This matrix Q has values $q_{ij} = \Sigma_l x_{ijl}c_{jl1}$, i.e. we replace the labels in P by the quantification of these labels on dimension 1.

The relation between MCA and *principal components analysis* is that the matrix Q has the property that the average of the squared correlation between each of the columns of Q and the object scores on dimension 1 is maximized and equal to the first eigenvalue found by MCA. This objective is similar to the objective in PCA, where Q is fixed, and the eigenvalue is maximized as a function of the object scores (called component scores) only. The relation between MCA and *generalized canonical analysis* runs along the same

line, but now each of the columns of Q is viewed as an optimal linear combination of the columns of their corresponding part of the indicator matrix. The relation between MCA and *analysis of variance* is that it is tried to make the values in each row of Q as *similar* as possible while at the same time the averages of the rows (which correspond to object scores) are made as *dissimilar* as possible. This is done under the restriction that each column of Q has average 0 and variance 1. It comes to the same as maximizing the average of the ratio of the 'between variance' and the 'total variance' for each variable, hence the relation with analysis of variance. For more details on these approaches we refer to Tenenhaus & Young (1985) and the references given there. Here we only emphasize the importance of the quantified matrix Q in many interpretations of MCA, since it makes some of the missing data approaches more easily understandable.

3. Missing data procedures in multiple correspondence analysis

In section 2 we discussed MCA without missing data. We approached MCA as a CA of an indicator matrix. Now we can introduce most of the missing data procedures described so far in the literature by simply showing how the presence of missing value in the original data leads to the adjustment of the indicator matrix. Consider again our small example in table 1a, in which we have inserted three missing values into row 3 and 4, see table 2a. We proceed by discussing the missing data procedures separately.

3.1 *Missing passive.*

This is a very popular approach to missing data. The idea is simply that $x_{ijl} = 1$ if object i falls into category l of variable j, and $x_{ijl} = 0$ if not. This implies that if the information for object i on variable j is missing, $x_{ijl} = 0$ for each category l of variable j (see table 2b). As a second step the centroid of objects that are missing on variable j can be calculated 'passively' (i.e. it does not influence the other analysis results) as a point for 'missing' on variable j, hence the name 'missing passive' (Gifi, 1981). This approach is suggested in Benzecri et al. (1973, p.327), Hamrouni & Benzécri (1976), Nishisato (1980) and Gifi (1981) and the properties of this approach are studied in detail by Meulman (1982). It is the default option for missing data in the program HOMALS (van de Geer, 1985).

Notice that the row margins are not all equal to m, the number of variables (see table 2b). This affects many of the MCA properties. First of all the metric for the category points is not the identity metric anymore. *Chi-squared distances* (1a) and (2a) can still be calculated, but cannot reduced further to (1b) and (2b) since x_{i++} is not necessarily equal to m. For (1a) this has the somewhat strange result that, if $x_{i++} \neq x_{i'++}$, and both objects fall into the category l of variable j, the distance $\delta^2(i,i')$ increases. The *total inertia* is not simply (k/m) - 1 anymore, but larger. This follows from (3b): since in 'missing passive' $x_{i++} \leq m$, it follows that $1/x_{i++} \geq 1/m$, and hence the total inertia becomes larger. The general equations (4) and (5a) cannot be simplified since D_r is not proportional to I anymore. However, the *transition formulas* define MCA still in a very simple way: using normalization (R,C*), each categories is still in the centroid of the objects that fall into it,

and using normalization (R*,C), each object (also an object with missing values) is in the centroid of the categories it falls into. A drawback of this approach is that, since D_r is not proportional to the identity matrix anymore, the interpretations of MCA in terms of principal component analysis, generalized canonical analysis and analysis of variance do not hold anymore (see Meulman, 1982, for a proof).

This approach has the advantage of being very simple, and the transition formulas can still be interpreted in the same way. However, many other important MCA properties are lost. Missing passive resembles so-called "available case" methods (compare Little & Rubin, 1987, ch. 3) that are sometimes used for data analysis with missing values: the resemblance is that only the non-missing values play an 'active' role in the analysis.

3.2 *Missing passive modified margin*

This approach, proposed by Escofier (1981,1987) is concerned with the property of the indicator matrix for *missing passive* that the row margins x_{i++} are not all equal to m, the number of variables. The option 'missing passive modified margin' solves this by artificially using constant row margins 1/n in the MCA calculations. So, compared to (4), we can write this approach in terms of generalized CA decomposition (see Escofier, 1983; van der Heijden & de Leeuw, 1985)

$$X/x_{+++} = S_r 1 S_c + S_r R \Lambda C' S_c \qquad (6)$$

where X is the same matrix is in 'missing passive', S_c is a diagonal matrix with the elements x_{+jl}/x_{+++}, but S_r is a diagonal matrix with values 1/n. Notice that the sum of the elements in $X/x_{+++} = 1$ whereas it is nm/x_{+++} for $S_r 1 S_c$. The objective of fixing the row margins artificially is to obtain a solution in which many of the MCA properties still hold. So 'missing passive modified margin' is meant to remedy the weak points of 'missing passive'.

Like in MCA without missing data, but unlike 'missing passive', the mettric for the category points is the identity metric. *Chi-squared distances* between the rows have the property that, when $x_{ijl}=x_{i'jl}=1$, this does not increase $\delta^2(i,i')$, since artificially $x_{i++}=x_{i'++}$. The *total inertia* is equal to (1d) (up to a proportionality factor nm/x_{+++}, to

remedy the fact that the total number of observed values x_{+++} is not nm). *Transition formula* (4b) holds as usual, i.e. using normalization (R,C*) the categories are still in the averages of the objects that used them; however, (4a) has a slightly different form, since the row margins of X/x_{+++} are not equal to the row margins of $S_r 1 S_c$ (compare van der Heijden & de Leeuw, 1985). Another nice property of 'missing passive modified margins' is that the relation with *principal component analysis, generalized canonical analysis* and*analysis of variance* again holds when we fill in $q_{ij} = 0$ if object i is missing on variable j, and the maximized eigenvalue is corrected with a factor nm/x_{+++}. It falls beyond the scope of this paper to discuss this option in more detail, we refer instead to Escofier (1987).

We conclude that, compared to 'missing passive', 'missing passive modified margin' has many nice MCA properties that are lost in 'missing passive'. As for 'missing passive' only the non-missing scores play an active role in the analysis, and as such this approach resembles "available case methods" also. The approach is computationally more difficult to use since a special program for generalized CA is needed.

3.3 *Missing single*

In the option 'missing single' for each variable having missing values a single extra category is created for the missing objects (see table 2b). This option for the treatment of missing data is very popular. Since it comes down to an ordinary MCA, all formulas in section 2 can be used in a straightforward way. Equation (1d) shows that, by adding columns to the indicator matrix, the *total inertia* increases with 1/m for each extra column. This shows that the missing categories take part 'actively' in the solution (as compared to 'missing passive'). Problems might result from this, for example, when the number of missing objects for some variable is very small, then the new categories may dominate the solution (Nishisato, 1980). Notice that the increase of the inertia (and hence of the influence on the solution) is relatively larger if the original number of categories is smaller (see (1d)). So if all variables have few categories 'missing single' will have a greater influence on the solution than if all variables have many categories. As far as the *transition formulas* are concerned, we find that, as usual, categories are in the centroid of the objects that chose them (this also holds for the categories for missing); objects are

now in the average of all their categories, i.e. not only the categories of the variables on which they were not missing. In the *quantified data matrix* Q the objects missing on variable j all have identical scores, that may very well be extreme.

We conclude that this approach is very simple to apply. It might sometimes lead to unwanted results: for example, when we are not really interested in the missing values, we might very well end up with a solution that distinguishes persons having missing values from persons not having missing values. In terms of general missing data procedures, this approach is not comparable to "available case methods" like missing passive, but instead to methods where optimal values are filled in into the data matrix. Notice that all objects that are missing are inserted into a single extra category. This is useful when it is assumed that these objects have something in common. It is generally suggested to use 'missing single' when the mechanism behind the missing data is not random, i.e. has a specific meaning (see, a.o., Hamrouni & Benzécri 1976; Bastin, 1980, p.310; Meulman, 1982; Greenacre, 1984; Benali, 1985; Benali & Escofier, 1987), or when one is not sure about the mechanism behind the missing data (Nishisato, 1980).

3.4 *Missing multiple*

In 'missing multiple' for each missing value in the original data matrix a separate extra column is created (see table 2b). So when a variable has more than one missing value, it receives multiple categories. This option is discussed in Gifi (1981), and studied in detail by Meulman (1982).

As in missing single, missing multiple also provides us with a complete indicator matrix, and therefore all equations in section 2 can be applied in a straightforward way. Similar remarks as for 'missing single' can be made for the *chi-squared distances*. Here, when objects have one or more missing values, they will be placed very far from the other objects in full-dimensional space. A striking point is that the *total inertia* increases considerably when the number of missing categories is large. Therefore this option is likely to produce outliers, given the tendency of MCA to place categories with small marginal frequencies in the periphery of the solution. The *transition formulas* show that,

using (R,C*), the missing categories receive the same scores as the objects for which they are missing. For more details we refer to Meulman (1982).

We conclude that this option is easy to interpret, like 'missing single'. Conceptually it is attractive in the case that it is assumed that the missing objects are missing for different reasons. However, this option is likely to produce outliers, especially if there are objects that have more than one missing value.

3.5 *Missing insertion*

In 'missing insertion' objects having missing values are inserted in some way into one of the original categories. Nishisato (1980; see also Nishisato & Levine, 1975) discusses two possible options to insert, namely "insertion of most consistent responses" and "insertion of least consistent responses". In the former approach objects with missing values are assigned to categories in such a way that the first eigenvalue is *maximized*, in the latter approach objects are assigned to categories in such a way that the first eigenvalue is *minimized* (for more details, see Nishisato, 1980). Van Buuren (1988) applies the k-means algorithm to assign an object with missing values to categories in such a way that the total inertia is maximized. In these options we deal with complete indicator matrices, and hence the properties of MCA all hold. They require a relatively large computational effort. He discusses the use of the above techniques for the determination of the range of the first eigenvalue, and the loss of information due to missing data. His aim is to indicate emperically when an analysis of the data should be given up, which is the case when the gap between the best and the worst configurations generated is too large to ignore (Nishisato, pers. comm.).

In the program SPADN (Lebart et al., 1987) a procedure is used that is closely related to inserting: there objects in categories with very low marginal frequencies are inserted *at random* into the other categories. Inserting at random seems to give good results if the main interest is in the configuration of category points.

3.6 *Missing fuzzy average*

In this option we fill in proportions for the missing categories. Much attention has been given to the following approach: when variable j has k_j prespecified categories, and some object is missing on this variable, it will receive a value $1/k_j$ for each category. So its value is distributed uniformly over the categories. We call this *fuzzy* because fuzzy coding is the name given to indicator matrices with values between zero and one (see Greenacre, 1984; van Rijckevorsel, 1987). This option is discussed in Benzécri et al. (1973, p.310), Hamrouni & Benzécri (1976), and properties of this option are studied in Cazes (1977, republished in Bastin et al., 1980), and Greenacre (1984). Benzécri et al. (1973) already warned for the inherent danger in this approach that when some category has a very low marginal frequency, an object with missing information receives a relatively large score on this category using this approach. This objection seems so serious to us, that we will not consider this possibility any further, but instead go to its natural generalization, that we coin *missing fuzzy average*.

In 'missing fuzzy average' we fill in the marginal proportions for the cells corresponding with a missing value. So for the first variable in table 2a this is 4/8 2/8 2/8, since 4 out of 8 non-missing objects fall into 'a', and so on; for the second variable the proportions are 4/9 3/9 2/9 (see table 2b). We call this *average*, because the average distribution for the non-missing objects is filled in into the indicator matrix. This approach was suggested by Benzécri et al. (1973, p.327) and Hamrouni & Benzécri (1976), and can be found in modified form in Greenacre (1984). These proportions can also be found using a procedure for missing data in ordinary contingency tables, namely 'reconstitution of order zero' (see Tallur, 1973; Mutumbo, 1973; Nora, 1975; Greenacre, 1984; de Leeuw & van der Heijden, 1988). When we apply reconstitution of order zero to the cells of an indicator matrix that correspond to the missing elements in the matrix of objects by variables, this procedure will iteratively find values that are independent given the new margins. These independent values are the average proportions. For 'reconstitution of order zero' it is known that the cells for which independent values are fitted do not contribute to the inertia (compare de Leeuw & van der Heijden, 1988). We will now discuss some other properties.

Due to the fact that the row margins $x_{i++}=m$, some of the MCA properties are still intact. *Chi-squared distance* formulas apply in a straightforward way. The total inertia will be lower than (k/m) - 1: the cells in which we find the marginal proportions do not contribute to the total inertia since $x_{ijl} = x_{i++}x_{+jl}/x_{+++}$ (compare (3a)). In this sense missing fuzzy average provides us with the *lowest possible total inertia* when we try to find an indicator matrix starting from an ordinary object by variable matrix as in table 2a. As far as the *transition formulas* are concerned, they can be applied straightaway: using (R*,C), an object is in the average of the categories that are non-missing; however, using (R,C*), a category is not in the average of the non-missing objects only, since the missing objects also contribute somewhat. In the *quantified data matrix* Q we find for the missing values the score 0, since $\Sigma_l x_{ijl} c_{jl\alpha} = 0$ for these cells. This clarifies what we do when we fill in average proportions into the indicator matrix: in terms of the quantified data matrix is comes down to filling in the average value (namely, zero) for the missing values of each variable. The optimality properties of MCA in terms of *principal component analysis*, *generalized canonical analysis*, and the *analysis of variance* do not hold anymore, however, due to the fact that the indicator matrix contains proportions.

Notice that by filling in the average it is not tried to extract information from the fact that someone is missing. On the contrary, a solution is sought in which, in terms of the total inertia, the influence of the missing entries is completely eliminated. It resembles the approach used in the context of missing quantitative data, where sometimes means are imputed for the missing values (compare Little & Rubin, 1987).

3.8 *Missing fuzzy subgroup.*

Missing fuzzy subgroup is a natural extension of 'missing fuzzy average': now we fill in proportions that are found for some subgroup, for example, when an older male has a missing value on some category, we calculate the proportions on this variable for the non-missing older males, and fill these in for the missing older male. For the missing entries in table 2a this implies the following: for row 3 we fill in the distribution of objects having scored 'q' 'w': this distribution is 1/2 1/2 0. For row 4 we fill in the distribution aving scored 'v': this is 2/3 1/3 0 2/3 0 1/3 (see table 2b). Though not precisely identical, this approach is suggested in Greenacre (1984, p.155). In this vein it is also possible to use more sophisticated methods to estimate the missing data (compare Little & Rubin,

1987, ch.9 and 11). It is tried to find a best guess of the score of an object, given its scores on the other variables. So, again, it is not tried to extract information from the fact that someone has a missing value on some variable.

Chi-squared distances can again be simplified since $x_{i++}=x_{i'++}=m$. The *transition formulas* apply, however, compared to missing fuzzy average, an individual is not anymore in the average of his non-missing variables only, since we have made a best guess for his missing variable. The *total inertia* is smaller than (k/m) -1 (see Greenacre, 1984; van Rijckevorsel, 1987) but larger than the total inertia for 'missing fuzzy average'. In the *quantified data matrix* Q the cells that correspond with missing values have a score that is in the average of the scores that the objects in its subgroup have. As in 'missing fuzzy average', MCA is not optimal in terms of *principal component analysis, generalized canonical analysis* or *analysis of variance.*

3.9 *Concluding remarks*

Nishisato (1980) distinguishes approaches that extract information from the fact that some object has a missing score, and approaches that do not do this. Obviously, 'missing single' and 'missing multiple' are options that extract information from being missing by defining 'being missing' explicitly as alternatives. Approaches that do not try to extract information from missing values are the available case methods 'missing passive', 'missing passive modified margin', the fuzzy options 'missing fuzzy average', 'missing fuzzy subgroup', and 'missing insertion'. In 'missing passive' an object score is only based on the non-missing scores, and in this sense no information is extracted from the fact that it has scores missing. The same holds for 'missing passive modified margin' that can be considered as an option that tries to remedy some MCA properties that are lost in 'missing passive'. In 'missing fuzzy average' always the average zero is filled in, and this shows that this option does not extract information from the missing scores. In 'missing fuzzy subgroup' an average is filled in for a subgroup, and this average does depend only on the non-missing information, so in this case we also do not extract information from the fact someone is missing. 'Missing insertion' doesnot use the fact that someone is missing for the determination of the category into which one is inserted. So we conclude that in this respect the main difference is between 'missing single' and 'missing multiple' on the one hand, and the other options. We have the impression that if

the number of missing values is small, the methods that do not extract information from being missing will give very similar results. First of all, the procedures 'missing passive' and 'missing passive modified margin' are almost identical. Compared with the fuzzy procedures and the inserting procedures, the indicator matrices only differ somewhat for the objects having missing values.

We have only discussed the approaches to missing data that are most distinct. We have ommitted some approaches from our study. For example, a set of combinations is found in Greenacre (1984), who combines in various ways 'missing single' together with the two fuzzy options. We have chosen to discuss these options separately in their 'pure' form. In Bastin et al. (1980) it is suggested to use reconstitution of order zero (see 'missing fuzzy average') to end up with an appropriate choice of zeros and ones for the missing values, but this is not worked out. A last option, discussed in Greenacre (1984) and Benali (1985) is to create one extra column, indicating the number of missing values each object has. This has the objective to have again an indicator matrix with constant row margins. Compared to missing single, this comes to the same, geometrically, as merging all missing categories into one point, their centroid (Greenacre, 1984).

4. Types of missing data

Here we will discuss distinct types of missing data that can appear in object by variable data (compare Little & Rubin, 1987). A first distinction is between data values that are really missing, and data values that are not really missing. We speak of data values that are really missing if in reality an object falls into one of the categories but we do not know into which category. As an example of data values that are not really missing, we might think of attitude items to which a person doesn't know the answer. In this case such a person cannot be classified in any of the prescribed categories . In the sequel we will refer to this type of missing data as *missing not categorizable*, abbreviated to MNC.

For data values that are really missing, Little & Rubin (1987) distinguish different classes of missing values, namely *missing completely at random* (MCAR), *missing at random* (MAR), and *missing not at random* (MNAR). Consider a categorical variable 'Salary' with one or more missing values, and the categorical variable 'Education' without missing. If the mechanism to be missing on 'Salary' does not depend on 'Salary' nor on 'Education', we speak of "missing completely at random". In this case the observed contingency table Salary x Education should be approximately equal to the not observed (due to the missing values on Salary) contingency table. If the mechanism to be missing on Salary does depend on Education but not on Salary, we speak of "missing at random". This implies that the conditional distribution of Salary given Education for the complete data can be used to estimate the missing values on Salary. If the mechanism to be missing on Salary depends on Salary and on Education, we speak of "missing not at random". This may be the case if the observed salary distribution of the higher educated persons differs from the non-observed salary distribution for the higher-educated persons that refused to specify their salary.

Another instance in which we deal with missing data in MCA is when we create them on purpose. We coin these missing data *created missing* (CM). The general idea of creating missing data is that we do not want certain categories to have an effect on the solution; therefore we define objects falling into these categories as missing. For example, we might eliminate categories with very low frequencies because they have produced outliers or they have dominated the solution in an earlier analysis. Another reason is that we consider objects in a specific class to be in that class for heterogeneous reasons, for

example, some persons do not agree with an item because it is to extreme whereas other do not agree because it is not extreme enough. Typical examples of this last phenomenon are 'pick any out of m'-data, seriation data, or voting data. For example, for 'pick any out of m'-data it is possible that only persons that pick some item are thought to be identical, since the persons who did not pick the item might have done so for different reasons. For seriation data, objects found in a tomb might indicate an identical period, whereas objects not found might indicate a period before or after the period in which these objects were found. And lastly, voters for some law are likely to have similar objectives, whereas voters against some law might do so for opposite reasons. For a thorough discussion of MCA of this type of data, we refer to Hamrouni & Benzécri (1976), Heiser (1981) and Meulman (1982).

We will now try to draw some conclusions from the properties of the different options for the treatment of missing data, by relating these properties to the different types of missing data. Our conclusions are summarized in table 3, having types of missing data in the columns, and the options for treating them in the rows. These conclusions will be helpful if we want to make a justified choice for any of the different options, given that we deal with a specific type of missing data. We will discuss table 3 columnwise.

4.1 *Missing not categorizable*

If an object having a missing value is not classifiable in the other categories, this object defines a new class. There are two possibilities: either we are interested in the relation of being missing with the other categories, or we are not interested in this relation.We assume here that in first instance we are interested in this relation. For the case that we are not interested in this relation we refer to section 4.5 where we discuss 'created missing' data.

Given that we are interested in being missing, we should extract information from it, and hence we can only use 'missing single' or 'missing multiple' (see section 3.9). In the other options the state of being missing does not play an active role in the analysis. 'Missing single' is to be preferred if we assume that those objects missing have something in common, for example, some persons ommit a question about children because it is irrelevant for them since they have no children. 'Missing multiple' is to be

preferred if it is assumed that the objects do not have something in common, for example, they ommit a question for different reasons. However, this option is likely to produce outliers.

4.2 *Missing not at random*

If we deal with data that are missing not at random, most often we do also best to extract information from the fact that someone is missing. In this way we can study, firstly, in which way 'being missing' is related to the categories of the variable for which the objects are missing, and, secondly, in which way 'being missing' is related to the categories of the other variables. As for 'missing not categorizable' we conclude therefore that we should use 'missing single' or 'missing multiple', depending on the fact whether we assume that those objects missing have something in common or not. In these options the fact that one is missing plays an active role in the analysis. Contrary to our discussion of 'missing not categorizable', we do not dismiss 'missing insertion' because it is a possible way to insert the missing objects into the category they would have had if they were not missing.

4.3 *Missing completely at random*

If we deal with data that are missing completely at random, it is perhaps best to *not* to use the option 'missing single' since it treats all objects missing in the same way whereas they do not have anything in common, due to the definition of 'missing completely at random'. On the other hand, 'missing multiple' seems elegant since is gives a category point for each missing score, but on the other hand it is very susceptible to producing outliers. Good choices are the options 'missing passive' and 'missing passive modified margin', since they only use the available information for each object. Using the fuzzy options can also be defended, since this corresponds to imputing the average for 'missing fuzzy average' or subgroup averages for 'missing fuzzy subgroup'. Obviously, the latter of these two fuzzy options is more elegant. 'Missing insertion' will also be a reasonable option in this context, since it can lead to inserting an object into its original category.

Our conclusions correspond roughly with findings from stability studies performed by Chan (1978; see also Nishisato, 1980, for a summary), Meulman (1982) and Benali (1985). In these studies data sets are used with a known structure, and in these data sets cells are defined completely at random as missing. It is evaluated how well the various options recover the properties of the complete data from the analysis of the incomplete data.

Chan (1978) uses a three-parameter logistic latent trait model for binary items. The generated data are one-dimensional. She compares, among others, 'missing passive' and 'missing single', and finds that the recovery of information in the complete data from the incomplete data declines rapidly for 'missing single' when the proportion of missing goes from 15 to 20%. On the other hand 'missing passive' works relatively good, even with 25% missing values.

Meulman (1982) compares the options 'missing passive', 'missing single' and 'missing multiple' using many criteria. Missing data are randomly inserted into three types of data sets, namely an algebraic data set, data generated under the multinormal distribution, and a set of real data that we do not discuss here further. All data sets are one-dimensional. The first data set has a rather strong first dimension. In general 'missing single' performs worse than 'missing passive' and 'missing multiple' when the number of missing values increases from 7 via 14 to 20%. There are two data sets generated with the multinormal distribution. The first has a rather weak first dimension: r=.30 for seven variables each discretized into five categories, with n=80. For 9% missing, 'missing passive' performs best; for 16% missing, 'missing passive' still performs best, though 'missing single' still performs acceptable, and 'missing multiple' is not acceptable anymore. The second data set is drawn from a multinormal distribution with r=.20, for n=1500; 14% missing values are imputed. In this case 'missing single' performs slightly better than 'missing passive', whereas 'missing multiple' should be avoided. Her main conclusion is that, when the data set has a strong first dimension, all three options behave rather well; when the first dimension is less strong, then missing passive and missing single are superior.

Benali (1985) compares 'missing passive modified margin' with at random inserting the objects missing in any of the categories. He uses real data that are clearly more-

dimensional in which he defined missing values completely at random. He finds that 'missing passive modified margin' gives somewhat more stable results than the alternative option. However, the evidence provided by this study is relatively small, since only one real data sets are used.

Important as these studies may be, it should be noticed that it only describes the behavior of some of the options under completely random missing data, and it concentrates mainly on the behavior of options in one-dimensional data sets. 'Missing passive' and 'missing passive modified margin' seem to perform adequately. Contradictory results are found by Chan and Meulman for 'missing single': Meulman explains the acceptable performance of 'missing single' in her study from the fact the data generated by her have more categories, and she concludes "the smaller the number of categories, the worse missing single is expected to behave" (Meulman, 1982, p.164). The reason for this is that, with a smaller number of categories, the relative increase of the total inertia is larger (see section 3.3).

4.4 *Missing at random*

Given the definition of 'missing at random' the option 'missing fuzzy subgroup' seems a very elegant alternative since it specifies probabilities that an object falls into any of the categories given its other scores. For the same reason 'missing fuzzy average', and 'missing insertion' are suboptimal options. We could use 'missing passive' and 'missing passive modified margin' if we only want to make use of the scores an object has. Conceptually, 'missing multiple' is also useful, but probably produces outliers. 'Missing single' is advized against since the objects missing do not necessarily stem from the same (unknown) class.

4.5 *Created missing*

Lastly, we might be dealing with data made missing on purpose because we want to eliminate the influence of some category. In this case we could best use 'missing fuzzy average', since this option minimizes the total inertia, and thus eliminates the influence of

this category completely. Other possible candidates are 'missing passive' and 'missing passive modified margin', in which no attempt is made to extract information from the fact that someone is missing. A drawback of 'missing fuzzy average' is that it does not work for binary variables: if one of both categories is eliminated, the average proportion for the other category becomes '1', and we have a column with all values equal to '1'.

5. Conclusion

In this paper we have discussed the ways for treating missing data that were most important in our opinion. We related these options for the treatment of missing data to the types of missing data that can occur. Table 3 indicates which option for the treatment of missing data will most often be preferable, given that one knows what type of missing data one is dealing with. We agrre with Nishisato (1980, see also section 3.3) that if the reason for being missing is unclear, it seems wise to assume that the data missing not at random and start with 'missing single'. Thus one is able to study the behavior of the extra categories.

For most of the types of missing data we pointed out that often, in our opinion, more than one option for the handling of missing data was acceptable. We never preferred 'missing multiple' due to the fact that is likely to produce outliers. 'Missing insertion' is never preferred because we think that it is only useful in very specific circumstances: in ordinary circumstances there are always better candidates, in our opinion. When more than one option is preferable, a choice between these options should be made if there are specific ideas about the properties of MCA that are thought to be important.

5. References

Bastin, Ch., Benzécri, J.-P., Bourgarit, Ch., Cazes, P. (1980). *Pratique de l'analyse des données. II: Abrégé théorique études de cas modèle*. Paris: Dunod.

Benali, H. (1985). *Stabilité de l'analyse en composantes principales et de l'analyse des correspondances multiples en présence de certain types de perturbation - méthodes de dépouillement d'enquêtes*. University of Rennes: Thèse de 3e cycle.

Benali, H. & Escofier, B. (1987). Stabilité de l'analyse factorielle des correspondances multiples en cas de données manquantes et de modalités à faibles effectifs. *Revue de Statistiques Appliquée, 35*, 41-52.

Benzécri, J.-P. and coworkers (1973). *L'analyse des données. Tome II: L'analyse des correspondances*. Paris: Dunod.

Carroll, J.D., Green, P.E., & Schaffer, C.M. (1986). Interpoint distance comparisons in correspondence analysis. *Journal of Marketing Research, 23*, 271-280.

Cazes, P. (1977). Problème sur l'analyse des questionnaires. *Cahiers de l'Analyse des Données, 1*, 73-77.

Chan, D.O. (1978). *Treatment of missing data by optimal scaling* Masters thesis, University of Toronto.

de Leeuw, J. & van der Heijden, P.G.M. (1988). Correspondence analysis of incomplete tables. *Psychometrika.*

Escofier, B. (1983). Analyse de la differance entre deux mésures sur le produit de deux memes ensembles. *Cahiers de l'Analyse des Données, 8*, 325-329.

Escofier, B. (1981, 1987). Traitement des questionaires avec non-response, analyse des correspondances avec marge modifiée et analyse multicanonique avec contrainte. *Publications de l'Institut de Statistique de l'Universite de Paris, 32*, 33-69 (published in 1981 as an internal report).

Gifi, A. (1981). *Non-linear multivariate analysis*. Leiden: D.S.W.O.-Press.

Greenacre, M.J. (1984). *The theory and applications of correspondence analysis*. New York: Academic Press.

Hamrouni, A. & Benzécri, J.-P. (1976). Les scrutins de 1967 à l'assemblée des nations unies. *Cahiers de l'Analyse des Données, 1*, 161-195, 259-286.

Heiser, W.J. (1981). *Unfolding analysis of proximity data*. Unpublished thesis, University of Leiden.

Lebart, L., Morineau, A., & Lambert, T. (1987). *SPADN = Système portable d'analyse des données*. Paris: CISIA.

Lebart, L., Morineau, A., & Warwick, K.M. (1984). *Multivariate descriptive statistical analysis.* New York: Wiley.

Little, R.J.A.,& Rubin, D.B. (1987). *Statistical analysis with missing data.* New York: Wiley.

Meulman, J. (1982). *Homogeneity analysis of incomplete data.* Leiden: D.S.W.O.-Press.

Meulman, J. (1986). *A distance approach to nonlinear multivariate analysis.* Leiden: D.S.W.O.-Press.

Mutumbo, F.K. (1973). *Traitement des données manquantes et rationalisation d'un reseau de stations de mésures.* Thèse de 3e cycle, Paris.

Nishisato, S. (1980). *Analysis of categorical data: dual scaling and its applications.* Toronto: University of Toronto Press.

Nishisato, S. & Levine, R. (1975). *Optimal scaling of omitted responses.* Paper presented at the Spring meeting of the Psychometric Society, Iowa City.

Nora, C. (1974). *Une méthode de réconstitution et d'analyse de données complètes.* Thèse de 3e cycle, Paris.

Tallur, B. (1973) *Analyse des correspondances en cas de données manquantes - Applications en biologie.* Thèse de 3e cycle, Paris 6.

Tenenhaus, M. & Young, F.W. (1985). An analysis and synthesis of multiple correspondence analysis, optimal scaling, dual scaling, homogeneity analysis and other methods for quantifying categorical multivariate data. *Psychometrika, 50,* 91-119.

van Buuren, S. (1988). *A procedure to estimate missing values in categorical data.* Paper presented at the meeting of the Psychometric Society, Los Angeles.

van de Geer, J.P. (1985). *HOMALS.* Leiden: Department of Data Theory.

van der Heijden, P.G.M., & de Leeuw, J. (1985). Correspondence analysis used complementary to loglinear analysis. *Psychometrika, 50,* 429-447.

van Rijckevorsel, J.J.A. (1987). *On horseshoes and fuzzy coding in multiple correspondence analysis.* Leiden: D.W.S.O-Press.

Table 1. An object by variable matrix with its indicator matrix

Table 1a. An object by variable matrix

a	p	v
b	q	w
b	q	w
a	r	v
c	r	w
b	p	v
a	p	w
c	p	w
a	q	w
a	r	v

Table 1b. The corresponding indicator matrix

a	b	c	p	q	r	v	w
1	0	0	1	0	0	1	0
0	1	0	0	1	0	0	1
0	1	0	0	1	0	0	1
1	0	0	0	0	1	1	0
0	0	1	0	0	1	0	1
0	1	0	1	0	0	1	0
1	0	0	1	0	0	0	1
0	0	1	1	0	0	0	1
1	0	0	0	1	0	0	1
1	0	0	0	0	1	1	0

Table 2: Row 3 and 4 of table 1a with missing values, and the corresponding indicator matrices

Table 2a. Row 3 and 4 of table 1a with missing entries on variable 1 and 2

?	q	w
?	?	v

Table 2b: Rows 3 and 4 of indicator matrices, illustrating distinct options to deal with missing entries

	a	b	c	?	?	p	q	r	?	v	w
Missing passive	0	0	0			0	1	0		0	1
	0	0	0			0	0	0		1	0
Missing single	0	0	0	1		0	1	0	0	0	1
	0	0	0	1		0	0	0	1	1	0
Missing multiple	0	0	0	1	0	0	1	0	0	0	1
	0	0	0	0	1	0	0	0	1	1	0
Missing fuzzy average	1/2	1/4	1/4			0	1	0		0	1
	1/2	1/4	1/4			4/9	3/9	2/9		1	0
Missing fuzzy subgroup	1/2	1/2	0			0	1	0		0	1
	2/3	1/3	0			2/3	0	1/3		1	0

Table 3: Types of missing data (columns) with options for their treatment (rows). The entries of the table indicate whether a combination can reasonably defended or not. Options that seem preferable are in bold. Abbreviations are MNC for 'missing not categorizable', MNAR for 'missing not at random', MCAR for 'missing completely at random', MAR for 'missing at random' and CM for 'created missing'. For details, see text.

	MNC	MNAR	MCAR	MAR	CM	
Missing passive (modified margin)	no	no	**yes**	**yes**	**yes**	Only zeroes Available case method
Missing single	**yes**	**yes**	no	no	no	One extra column per var. active role
Missing multiple	yes	yes	yes	yes	no	One extra col. per miss.val. active role outliers
Missing insertion	no	yes	yes	yes	no	Inserted in original cat.
Missing fuzzy average	no	no	yes	yes	**yes**	Mean imputed, Inertia minimal
Missing fuzzy subgroup	no	no	**yes**	**yes**	no	Subgroup mean imputed

LISTE DES DERNIERES PUBLICATIONS INTERNES

PI 416 - CONTROLE D'EXECUTION EN ROBOTIQUE DE COOPERATION
Catherine JAUFFRINEAU
52 Pages, Juin 1988.

PI 417 - UN SCHEMA (ABSTRAIT) D'ITERATION REPARTIE. APPLICATION AU CALCUL DES CHEMINS DE VALEURS MINIMALES
Jean-Michel HELARY, Michel RAYNAL
24 Pages, Juin 1988.

PI 418 - STEREOMETRIE INTERACTIVE POUR LA TELEOPERATION ASSISTEE PAR ORDINATEUR
Alain VERDIER
52 Pages, Juin 1988.

PI 419 - PARTAGE D'OBJETS DANS LES SYSTEMES DISTRIBUES. PRINCIPES DES RAMASSE-MIETTES
André COUVERT, Aomar MADDI, René PEDRONO
60 Pages, Juin 1988.

PI 420 - NONCONFORMING FINITE ELEMENTS FOR THE STOKES PROBLEM
Michel CROUZEIX, Richard S. FALK
26 Pages, Juillet 1988.

PI 421 - COMPLING TEMPORAL LOGIC SPECIFICATIONS INTO OBSERVERS
Omar DRISSI-KAITOUNI, Claude JARD
18 Pages, Juillet 1988.

PI 422 - DISTANCE MEASURES FOR SIGNAL PROCESSING AND PATTERN RECOGNITION
Michèle BASSEVILLE
50 Pages, Septembre 1988.

PI 423 - MULTIPLE CORRESPONDENCE ANALYSIS WITH MISSING DATA
Peter G.M. van der HEIDJEN, Brigitte ESCOFIER
30 Pages, Septembre 1988.

PI 424 - CLOSED FORM SOLUTION FOR THE DISTRIBUTION OF THE TOTAL TIME SPENT IN A SUBSET OF STATES OF A HOMOGENEOUS MARKOV PROCESS DURING A FINITE OBSERVATION PERIOD
Bruno SERICOLA
20 Pages, Septembre 1988.

Imprimé en France
par
l'Institut National de Recherche en Informatique et en Automatique